When 0 is good: instrumental learning with counterintuitive goals decreases working memory engagement

Ti-Fen Pan (tfpan@berkeley.edu)
Department of Psychology,
University of California, Berkeley,
Berkeley, CA 94720 USA

Gaia Molinaro (gaiamolinaro@berkeley.edu)
Department of Psychology,
University of California, Berkeley,
Berkeley, CA 94720 USA

Anne GE Collins (annecollins@berkeley.edu)
Department of Psychology, Helen Wills Neuroscience,
University of California, Berkeley, Berkeley, CA 94720 USA

Abstract

Humans are adept at setting goals quickly and flexibly in their daily lives. Previous research has shown that people can assign rewarding properties to abstract or novel outcomes and use them to guide behavior. However, the mechanisms supporting this flexibility and their impact on learning processes, such as working memory (WM) or slower incremental systems, remain unclear. To address this, we designed an instrumental learning task in which participants learned stimulus-action associations by pursuing either standard goals (+1) or counterintuitive goals (+0) under varying WM loads. Our behavioral and modeling results revealed that when pursuing counterintuitive goals, humans learned more slowly and shifted their reliance from WM to habit-like associative processes, despite both processes remaining functionally intact. Additionally, we replicated previous findings showing that humans do not rely on reinforcement learning (RL) processes but instead integrate WM and habit-like processes to learn the associations. This interplay between WM and habit-like processes may allow a more resource-efficient approach to pursuing diverse goals. Our findings shed light on the breadth and cost of people's ability to flexibly learn and pursue any goal.

Keywords: goal-directed learning; human reinforcement learning; working memory; computational cognitive modeling

Introduction

Humans can flexibly assign rewarding properties to various goal outcomes to guide behavior (O'Reilly, 2020; De Martino & Cortese, 2023). For example, X-crosses are often associated with negativity or incorrectness, but in tic-tac-toe, players effortlessly adopt X-crosses as winning elements to form a line. This dynamic goal reward property-setting enables us to adapt to diverse contexts. Classic theories of human reinforcement learning, however, assume that rewards are inherently tied to environmental features or serve as proxies for these features (e.g., secondary reinforcers). By contrast, we propose that biological agents internally and flexibly adjust the rewarding strength of external signals based on multiple factors such as current goals.

Recent evidence highlights the primacy of goal-congruent outcomes over subjective reward value (Frömer, Dean Wolf, & Shenhav, 2019). For example, Frömer et al. (2019) demonstrated that the brain's value network exhibits strong responses to goal-congruent outcomes, even when these outcomes conflict with subjective rewards. Similarly, fMRI studies revealed that individuals can pursue abstract, non-valenced goals by engaging the brain's reward circuitry to support learning (McDougle, Ballard, Baribault, Bishop, & Collins, 2022). Despite this flexibility, pursuing abstract or neutral goals is often associated with diminished learning performance, though the cognitive mechanisms behind this decline remain poorly understood (Molinaro & Collins, 2023a).

This study builds on prior work (McDougle et al., 2022; Molinaro & Collins, 2023a) by extending the investigation of associative, goal-dependent learning to counterintuitive goals – those associated with strong negative valence or non-rewarding priors. Such goals pose a greater challenge to the human goal-dependent reward-setting process and allow us to probe its adaptability under less conventional conditions. Unlike studies that present feedback involving binary outcomes (e.g., +1/+0 or +0/-1) to compare reward seeking vs. loss avoidance (Palminteri et al., 2012), our experiment explicitly instructed participants to pursue specific goal values (+1 or +0). This setup focuses on reward-based learning by requiring participants to treat counterintuitive outcomes (e.g., 0-point outcomes) as desirable goals.

Furthermore, learning from feedback is supported by multiple cognitive processes, including working memory and habit-like mechanisms, which integrate outcomes in distinct ways (A. G. Collins & Frank, 2012). To disentangle these contributions, we employ variants of the RLWM task (A. G. Collins, Brown, Gold, Waltz, & Frank, 2014), which are specifically designed to recruit these processes to varying degrees. This approach allows us to explore cognitive processes that support learning from counterintuitive goals, shedding light on the limits and flexibility of human goal-reward adaptation.

Methods

Experimental design

We adapted the RLWM task (A. G. Collins et al., 2014) to disentangle the contributions of working memory (WM) from those of slower cognitive processes to instrumental learning by manipulating information load. On each trial, participants were first informed of which outcome value (+1 or +0) was their 'GOAL' outcome; then, they were asked to respond to a stimulus with one of three key presses, followed by outcome feedback (Fig 1 A). Participants were instructed to "collect" goal outcomes using a separate key press. Over trials, they learned the unique correct response associated with each specific stimulus, based on deterministic goal-contingent feed-

In D. Barner, N.R. Bramley, A. Ruggeri and C.M. Walker (Eds.), *Proceedings of the 47th Annual Conference of the Cognitive Science Society*

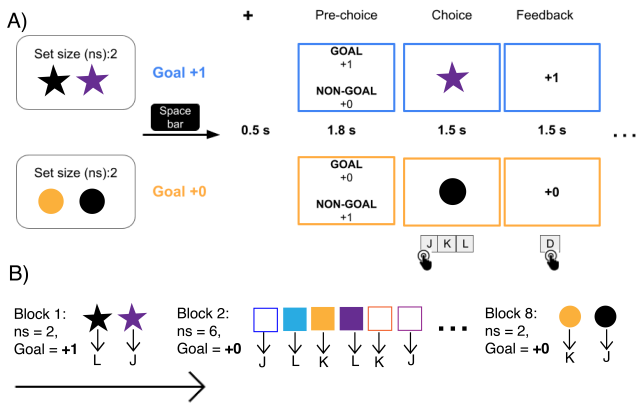

Figure 1: **Experimental paradigm**. A) An example of a set size 2 block trial in standard (+1 as a goal, here highlighted in blue) and counterintuitive (+0 as a goal, in yellow) conditions. B) Participants learned the association between stimuli (novel images) and actions (key presses) in each block from deterministic goal outcomes.

back. Each block had a set size (i.e., number of stimulus-response associations to learn) of 2 or 6, and either +1 or +0 Goal condition. Each stimulus was presented 13 times, pseudo-randomly interleaved. There were two blocks for each combination of set size and Goal condition, resulting in a total of eight blocks, with new images used in each block.

Participants A total of 118 participants completed the experiment online: 55 were recruited online from the undergraduate participant pool at the University of California, Berkeley, and 63 were recruited through Prolific. All participants provided informed consent online. Undergraduates received partial course credit as compensation, while participants recruited through Prolific were compensated at a rate of $12 per hour.

Twenty-eight participants were excluded based on the following criteria: 1) missing more than 25 trials during the experiment, 2) selecting the same response more than 20 consecutive times, 3) having an average reward collection error rate exceeding 20% in Goal 1 and 35% in Goal 0, 4) having an average response time in choice and reward collection faster than 300 ms. This resulted in a final effective sample size of 90 participants (age in years = 24.3 ± 6.3; 66 females) for our analysis. The experimental protocol was approved by the Institutional Review Board at the University of California, Berkeley.

Computational Modeling

We used a mixture modeling framework (A. G. Collins, 2018; A. Collins, 2024) to investigate multiple cognitive processes engaging in our instrumental learning task (Fig. 1). The mixture modeling can be broken down into two major components: a working memory (WM) process to capture fast but forgetful information integration, and a slow but non-forgetful integrative process. The policy π_{mixture} integrating these two processes is:

$$\pi_{\text{mixture}}(a|s) = \rho_{\text{WM}}(ns)\pi_{\text{WM}}(a|s) + (1-\rho_{\text{WM}}(ns))\pi_{\text{other}}(a|s)$$

where a and s denote the performed action and the presented stimulus. The WM weight $\rho_{\text{WM}}(ns))$ captures the engagement of the WM process and is parameterized per set size: $\rho_{\text{WM}}(ns=2)$ and $\rho_{\text{WM}}(ns=6)$. π_{other} represents the policy from either an RL or a habit-like (H) process.

A uniform random policy is added to the final policy and captures random lapses in choices, with a noise parameter $\varepsilon \in [0,1]$:

$$\pi_{\text{mixture}} \leftarrow (1-\varepsilon)\pi_{\text{mixture}} + \varepsilon\frac{1}{n_A}$$

where n_A denotes the total number of possible actions in a task (three in our experiment Fig. 1 A).

WM module The WM module tracks stimulus-action associations after observing stimuli, actions, and rewards (s_t, a_t, r_t) at trial t. The strength of each association is initialized as $W_0 = \frac{1}{n_A}$ and updated as:

$$W_{t+1}(s_t,a_t) = W_t(s_t,a_t) + \alpha_{\text{WM}}(r_t)\left(r_t - W_t(s_t,a_t)\right),$$
$$\text{where } \alpha_{\text{WM}}(r_t) = \begin{cases} 1, & \text{if } r_t = 1, \\ \text{bias}_{\text{WM}}, & \text{if } r_t = 0. \end{cases}$$

where we applied one-shot encoding for positive outcomes ($r_t = 1$), while using a $\text{bias}_{\text{WM}} \in [0,1]$ parameter to capture a potential neglect of negative outcomes ($r_t = 0$). The WM module updates stimulus-action association strengths immediately after experiencing outcomes but is susceptible to memory decay:

$$\forall (s,a), W_{t+1}(s,a) = W_t(s,a) + \phi_{\text{WM}}\left(W_0 - W_t(s,a)\right)$$

where $0 \leq \phi_{\text{WM}} \leq 1$ is a decay rate parameter. The WM module produces the final action based on the weights transformed by a standard softmax β:

$$\pi_{\text{WM}}(a|s) = \frac{\exp\left(\beta W(s,a)\right)}{\sum_i \exp\left(\beta W(s,a_i)\right)}$$

where the inverse temperature parameter β is fixed to 25 in our modeling to improve parameter recovery reliability (Wilson & Collins, 2019; Senta, Bishop, & Collins, 2025).

RL module The RL module tracks each stimulus-action association Q through a classic delta-rule model:

$$Q_{t+1}(s_t,a_t) = Q_t(s_t,a_t) + \alpha_{\text{RL}}(r_t)\left(r_t - Q_t(s_t,a_t)\right),$$
$$\text{where } \alpha_{\text{RL}}(r_t) = \begin{cases} \alpha, & \text{if } r_t = 1, \\ \alpha * \text{bias}_{\text{RL}}, & \text{if } r_t = 0. \end{cases}$$

Q_0 is initialized as $\frac{1}{n_A}$. The positive learning rate $\alpha \in [0,1]$ is a free parameter, and the negative learning rate is parameterized by a $\text{bias}_{\text{RL}} \in [0,1]$ parameter, which can be free or fixed depending on the specific model. The RL module employs the same softmax policy as the WM module with a fixed $\beta = 25$:

$$\pi_{\text{RL}}(a|s) = \frac{\exp(\beta Q(s,a))}{\sum_i \exp(\beta Q(s,a_i))}$$

H module The H module is similar to the RL module, with one key distinction: the subjective outcome (*SR*) is fixed at 1. This design reflects the agent's insensitivity to feedback, focusing solely on stimulus-action associations (*H* values), akin to Hebbian learning (A. Collins, 2024):

$$H_{t+1}(s_t,a_t) = H_t(s_t,a_t) + \alpha_H(r_t)\left(\text{SR}(r_t) - H_t(s_t,a_t)\right),$$
$$\text{where } \alpha_{\text{H}}(r_t) = \begin{cases} \alpha, & \text{if } r_t = 1, \\ \alpha * \text{bias}, & \text{if } r_t = 0. \end{cases}$$

H_0 is initialized as $\frac{1}{n_A}$. Like RL, the H module selects actions through a softmax policy with a fixed $\beta = 25$:

$$\pi_{\text{H}}(a|s) = \frac{\exp(\beta H(s,a))}{\sum_i \exp(\beta H(s,a_i))}$$

Sticky choice We incorporated a shared "stickiness" parameter $\kappa \in [-1,1]$ across all modules to account for stimulus-independent choice perseveration – the tendency to repeat the same key press across consecutive trials regardless of the presented stimulus. This was integrated into the softmax policy as:

$$\pi_{\text{WM}}(a \mid s) = \frac{\exp(\beta W(s,a) + \kappa I(a,a_{t-1}))}{\sum_i \exp(\beta W(s,a_i) + \kappa I(a,a_{t-1}))},$$
$$\text{where } I(a,a_{t-1}) = \begin{cases} 1, & \text{if } a = a_{t-1}, \\ 0, & \text{if } a \neq a_{t-1}. \end{cases}$$

Model candidates We explored three models: WM+H ($bias_{WM} = bias_H$) , WM+RL0 (free $bias_{WM}$, $bias_{RL} = 0$), and WMRL ($bias_{WM} = bias_{RL}$). These models were selected based on findings from a previous study (A. Collins, 2024), where WM+H emerged as the best-performing model among all mixture models, while WM+RL0 was identified as the strongest candidate within the RL modules.

Model fitting All models were fitted using maximum likelihood estimation with Python's scipy library. The L-BFGS-B algorithm (Byrd, Lu, Nocedal, & Zhu, 1995) was applied for bound constrained minimization using eight random starting points and a maximum of 1,000 iterations per participant. Each candidate model was fitted separately for each goal condition. We further conducted parameter identifiability analyses on the winning model (Wilson & Collins, 2019). This involved two steps: (1) simulating data based on the best-fitting (true) parameters, and (2) recovering the parameters from the simulated data and assessing their correlation with the true parameters. We confirmed that all parameters were identifiable, as each Spearman correlation was statistically significant ($p < 1e-3$).

Model comparison We computed model frequencies and protected exceedance probabilities (Rigoux, Stephan, Friston, & Daunizeau, 2014) for each goal condition, using the Akaike Information Criterion (AIC) (Akaike, 1974) as the log evidence for the calculations. To validate the winning model, we simulated the model using the best-fitting parameters, generating five simulated agents per participant.

Results

Humans can adaptively learn from counterintuitive goal outcomes

Participants successfully learned stimulus-action associations across all set sizes and conditions. In set size 2 blocks, performance was strong in both Goal 1 (standard) and Goal 0 (counterintuitive) conditions, with an accuracy of 90% and 83% respectively ($t_1(90) = 77.1$, $t_0(90) = 42.4$, $p < 1e-8$ in both, chance level=33%). In set size 6 blocks, participants maintained good performance, achieving an accuracy of 75% in Goal 1 and 71% in Goal 0 conditions. After 10 stimulus iterations (Fig 2), participants achieved 94% accuracy in set size 2 and 85% in set size 6 under the Goal 0 condition ($t_{ns=2}(90) = 49.3$; $t_{ns=6}(90) = 26.7$, $p < 1e-8$ in both). These findings indicate that individuals can learn well even when learning from counterintuitive goal outcomes.

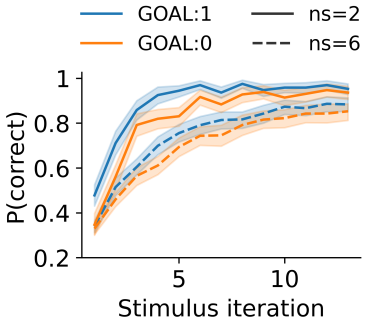

Figure 2: Participants' mean accuracy across stimulus iterations under different conditions. The shaded area indicates the 95% confidence interval (CI).

Learning efficiency decreases when pursuing counterintuitive goals

We observed lower average accuracy (learning performance) in the Goal 0 condition ($M = 77\% \pm 0.01$) compared to Goal 1 ($M = 83\% \pm 0.01$) across all set sizes (Fig 3 Left). In set size 2, post hoc tests revealed a significant main effect of goal condition (Goal 1 vs. Goal 0 repeated measures ANOVA: $F_{1,89} = 33.65$, $p < 1e-4$), with better learning performance in the Goal 1 (standard) condition. A similar effect was found in set size 6 ($F_{1,89} = 10.71$, $p = 0.002$). Set size (cognitive

load) also showed a significant main effect on accuracy (ns=2 vs. ns=6; $F_{1,89} = 114$, $p < 1e-4$), with larger set sizes (6) leading to lower accuracy compared to smaller set sizes (2).

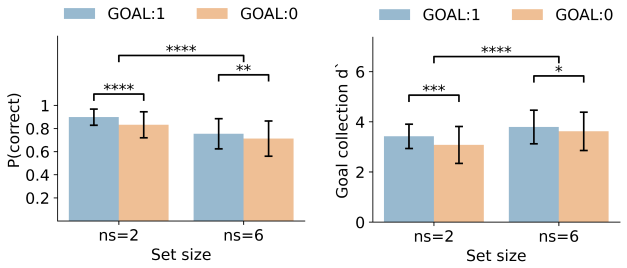

Figure 3: **Left.** Participants' mean accuracy as a function of goal conditions and set sizes. **Right.** Goal collection d'. Significant differences were found between the goal conditions (Goal 1 vs. Goal 0) and across set sizes (ns=2 vs. ns=6). Error bars indicate 95% CI. $^{*}p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$; $^{****}p < 0.0001$.

To assess whether these findings held true at the more granular level of stimulus iteration, we ran a mixed-effects linear regression analysis (Table 1). Confirming our ANOVA results, we find a significant main effect of goal type, where participants performed worse when pursuing the counterintuitive goal (+0) compared to the standard goal (+1). Additionally, set size had a significant effect, with higher cognitive (WM) loads (ns=6) leading to lower accuracy than lower loads (ns=2). There was no interaction between goal type and set size, suggesting that the detrimental effect of pursuing counterintuitive goals was consistent across cognitive loads.

Interestingly, a significant interaction between goal type and iteration (z=2.374, p=0.018) revealed that the impact of goal type on accuracy evolved over the course of learning during a block. This suggests that participants may have adapted to the counterintuitive goal (+0) over time, reducing the performance gap between the two goal conditions as they progressed through the block (See Fig. 2).

One potential explanation for the lower performance in the Goal 0 condition is that participants were more prone to lapses, misidentifying +1 outcomes as obtaining their goal. If this were the case, we would expect similar effects of experimental condition on goal collection accuracy, measured as d', as on learning performance. Indeed, d' was significantly lower in the Goal 0 condition compared to Goal 1 for both set sizes (Fig 3 Right). In set size 2, an ANOVA on d' showed a significant effect of feedback condition ($F_{1,89} = 16.3, p < 1e-3$), and a similar trend was observed in set size 6 ($F_{1,89} = 3.69, p = 0.05$).

However, a more detailed analysis using mixed-effects linear regression revealed a different pattern for goal collection d' compared to learning performance. While there was a significant negative main effect of goal type (Table 1) on goal collection, the effect of set size was in the opposite direction. Specifically, a positive coefficient for set size (0.336) indicated that smaller set sizes (2) were associated with *reduced* d'. A potential explanation for this surprising finding may be that when participants mastered stimulus-action associations in easier tasks, they paid less attention to goal outcomes. Yet, more generally, these results indicate that while lapses in goal recognition may play a role, they cannot fully explain the observed impairments in learning when pursuing counterintuitive goals.

Table 1: Mixed-effects linear regression predicting key press accuracy and goal collection d' as a function of the Goal type (+1 v.s. +0), set size, stimulus iteration, and their interactions.

Predictor	**Estimated ± SEM**	**z**	**p**
Accuracy			
Intercept	0.825 ± 0.02	41.7	***
Goal type	-0.098 ± 0.02	-4.9	***
Set size	-0.198 ± 0.02	-9.5	***
Iteration	0.138 ± 0.02	6.9	***
Goal type x iteration	0.067 ± 0.03	2.4	*
Set size x iteration	0.87 ± 0.03	3.1	**
Goal type x Set size	0.045 ± 0.03	1.6	0.11
Goal collection d'			
Intercept	3.377 ± 0.11	31.6	***
Goal type	-0.44 ± 0.13	-3.4	**
Set size	0.336 ± 0.13	2.6	*
Iteration	0.019 ± 0.13	0.2	0.88
Goal type x iteration	0.182 ± 0.18	1.0	0.32
Set size x iteration	-0.126 ± 0.18	-0.7	0.49
Goal type x Set size	0.301 ± 0.18	1.6	0.1

Note: *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$.

A habit-like process complements working memory to support goal-dependent behavior

To investigate what cognitive processes support goal-dependent learning, we tested three model candidates described in section Computational Modeling and fitted each model in Goal 1 and Goal 0 conditions, separately. Surprisingly, the inferred model frequency (Rigoux et al., 2014) (Fig. 4A), which represents the proportion of participants assigned to each model, showed that in both goal conditions, the WMH model combining a working memory (WM) process with a habit-like (H) agent outperforms the WM model combined with a reinforcement learning (RL) agent. To understand why WMH provided a better fit than WMRL, we conducted an error analysis following the approach in A. Collins (2024). For each participant error, we counted previous occurrences of the same error (chosen error, CE) and of other possible errors (unchosen error, UE) up to trial $t-1$, shown as blue and purple in Fig. 4B. Briefly, RL and WM models predict CE should be lower than UE (as we learn to avoid unrewarding actions), while H predicts UE should exceed CE.

WMH can therefore accommodate small or absent CE-UE differences, which RLWM models cannot. Our findings replicate recent results (A. Collins, 2024), showing that RLWM fails to capture the tendency to repeat the same stimulus-action errors at higher set sizes (Fig. 4B). By contrast, WMH explains this pattern via a habit-like mechanism that reinforces frequently repeated stimulus-action associations. We further validated the winning model WMH by simulating choice data using the fit model parameters. As shown in Fig. 4 C, the model is highly consistent with participants' learning curves across stimulus iterations. Taken together, our results suggest that humans use WM, a flexible executive function, to quickly adapt the goal outcomes while incorporating a simple, habit-like process (outcome-insensitive) to strengthen stimulus-action associations.

Working memory engagement decreases when pursuing counterintuitive goals compared to standard goals

Next, we attempted to identify which processes are responsible for weaker learning in the Goal 0 condition by testing how model parameters differed across goal conditions. A Wilcoxon test revealed that the $WM_{ns=2}$ weight in Goal 1 was significantly higher than in Goal 0 (Goal 1 vs Goal 0, $W = 535,\ p < 1e-3$) (Fig 4 D). The same pattern was found for $WM_{ns=6}$ weight ($W = 1476,\ p = 0.02$).

In contrast, no significant differences in the WM decay rate were observed between goal conditions ($W = 1505, p = 0.98$). This indicates that participants relied less on WM when pursuing counterintuitive goals, without any evidence of impairment to the WM process itself. If the WM process were impaired, we would expect a significantly lower WM decay rate for counterintuitive goals, as participants would be more prone to forgetting what they had learned. Similarly, we found no significant differences in the learning rate (α; $W = 1879, p = 0.49$), which governs habit-like learning. This suggests that the incremental habit process remained intact across both conditions (Fig 4D).

Discussion

In this study, we show that humans can flexibly assign rewarding properties to counterintuitive (non-rewarding) outcomes in an instrumental learning task. While participants eventually achieved similar performance across different goal types after several iterations, their average accuracy was consistently lower when pursuing 0-point outcomes as a goal, irrespective of working memory (WM) load. Replicating prior findings (A. Collins, 2024), our modeling results suggest that goal-dependent learning (in both standard and counterintuitive goal conditions) is supported by a fast WM process and a slower, habit-like (H) associative process, without relying on RL systems. Notably, the reduced WM engagement when pursuing counterintuitive outcomes explains the observed decreases in performance. This diminished WM involvement suggests additional executive control is needed to

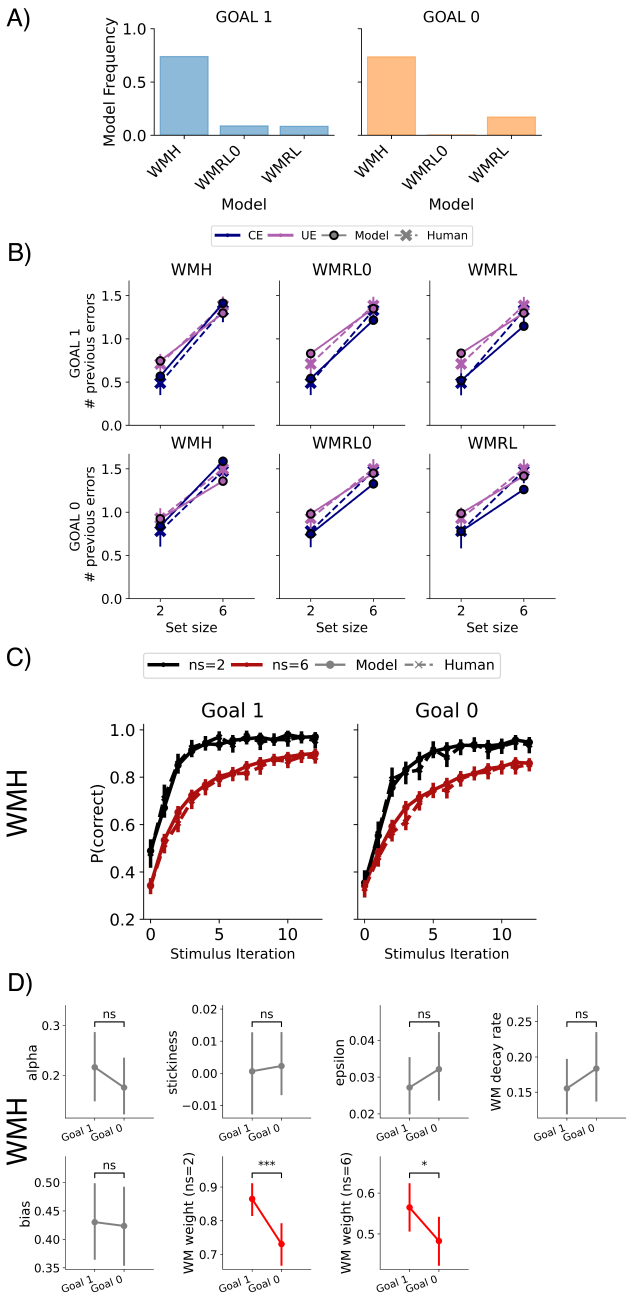

Figure 4: A) **Model comparison**: Each bar shows model frequency across participants. The winning model, WMH, has a protected exceedance probability of 0.99 in both conditions. B) **Error analysis**: Number of previoius errors by type, for both participants and model predictions. CE: chosen errors; UC: unchosen errors. Error bars indicate SEM across participants. Dashed lines:empirical data; solid lines:simulations. C) **Model validation**: Mean $\pm$ SEM across participants of participant data and WMH model simulation. Dashed lines:empirical data; solid lines:simulations. Black/red lines represent set sizes 2/6. D) **Model parameters**: Mean and 95% CI of WMH model parameters across goal conditions. $^{*}p < 0.05$; $^{***}p < 0.001$; Wilcoxon test

maintain counterintuitive goal representations, thereby limiting the cognitive resources available for WM. Additionally, the simpler H process, resembling Hebbian learning, may provide a more effective strategy for flexible goal pursuits because of its insensitivity to outcome values, in contrast to the reinforcement learning (RL) process. (Miller, Shenhav, & Ludvig, 2019; A. Collins, 2024).

Consistent with prior research on learning by pursuing abstract and neutral goals (McDougle et al., 2022; Molinaro & Collins, 2023a), our findings confirm that humans can adaptively pursue various outcomes, albeit with reduced efficiency. However, we observed a significant difference in goal collection accuracy across goal conditions, which deviates from the results reported by Molinaro and Collins (2023a) (i.e., comparing our results for $ns = 6$ with Experiment 2 in Molinaro and Collins). This discrepancy might be attributed to strong biases against counterintuitive goals, leading to more frequent errors in goal recognition compared to neutral outcomes. Nevertheless, the opposing main effect of set size on goal collection accuracy supports the conclusion reached in prior studies: lapses in goal recognition likely play only a marginal role in reducing learning efficiency.

One possible explanation for the observed performance differences is a potential difficulty in assigning reward values to counterintuitive goals stemming from prior negative associations. Our modeling did not explicitly account for subjective rewards, and it is plausible that individuals require more time to adjust their internal reward representations for counterintuitive goals. Future research could address this question by modeling subjective reward computations directly (Molinaro & Collins, 2023b). Another plausible explanation involves reduced motivation associated with counterintuitive goals. However, this should have translated to overall noisier behavior (e.g., as an increase in the noise parameter ε), which we did not observe (Fig. 4 D). Nonetheless, future work should further investigate these alternative explanations.

A key question is how humans assign rewarding properties to counterintuitive goals and use them to guide learning. Our modeling suggests that adaptation to such goals relies on cognitive processes sharing limited executive resources, evidenced by reduced – albeit unimpaired – WM engagement in these conditions. One explanation is that counterintuitive goals require additional cognitive control to update and maintain internal representations (e.g., perceiving a +0 outcome as desirable). This internal conflict taxes executive control, reducing resources available for WM processes. This aligns with neuroimaging evidence from previous research (McDougle et al., 2022), which underscores the role of executive function. Specifically, prefrontal cortex (PFC) activation during both the pre-choice and feedback phases when pursuing abstract goals in a two-choice probabilistic learning task was predictive of learning efficiency from abstract goals.

Although our study identifies key cognitive processes involved in flexible goal-dependent learning, exactly how these processes contribute to the assignment of rewarding properties remains unclear. Addressing this question requires a more integrated approach to modeling the mechanisms underlying goal adaptation. In our modeling analysis (Fig. 4), we fit the data for each goal type separately, which allowed us to isolate and compare the cognitive processes associated with standard and counterintuitive goals. However, this approach does not fully capture the shared and distinct mechanisms that enable the flexible assignment of reward properties across diverse goal types. Future research should explore fitting all goal types simultaneously under the same computational model. This approach could allow researchers to disentangle the contributions of shared versus goal-specific parameters, providing a clearer understanding of how cognitive processes, such as WM and habit-like mechanisms, interact to adaptively assign reward properties.

A key area for future exploration involves individual differences in WM capacity and their role in adapting to counterintuitive goals. While participants in our study relied on a combination of WM and habit-like processes to achieve counterintuitive goals, it is unclear whether individuals with greater WM capacity might exhibit enhanced adaptation. For instance, participants with superior WM resources may rely more heavily on WM processes and less on habit-like processes, potentially mitigating the performance decrement observed with counterintuitive goals.

Another avenue for investigation is the generalizability of our findings to other types of counterintuitive goals. This study focused on a specific counterintuitive goal (+0) in an instrumental learning context. However, future research could explore whether similar cognitive processes underlie adaptation to other types of counterintuitive goals, such as goals typically associated with negative outcomes (e.g., punishment avoidance). Understanding whether the observed reliance on habit-like processes and the decreased contribution of WM extend to these scenarios would provide a more comprehensive understanding of the flexibility and limitations of human goal adaptation. Such research could have profound real-world implications. For instance, dieting – pursuing the goal of consuming fewer calories – can be considered a counterintuitive objective. This highlights the importance of employing adaptive cognitive strategies when faced with such challenging goals. Our results suggest that successfully navigating such goals may involve strategic allocation of our limited executive resources.

Taken together, the behavioral results and computational modeling provide valuable insights into the cognitive processes underlying adaptation to counterintuitive goal pursuits. By uncovering the interaction between WM and habit-like processes, our study presents an alternative computational perspective on the mechanisms of instrumental learning. It highlights how reduced WM engagement may lead to lower performance when pursuing counterintuitive outcomes, while the habit-like process plays a complementary role in supporting adaptation to diverse goal pursuits.

Acknowledgments

This work was supported by NSF Grant 2336466 (Collins).

References

Akaike, H. (1974). A new look at the statistical model identification. *IEEE transactions on automatic control*, *19*(6), 716–723.

Byrd, R. H., Lu, P., Nocedal, J., & Zhu, C. (1995). A limited memory algorithm for bound constrained optimization. *SIAM Journal on scientific computing*, *16*(5), 1190–1208.

Collins, A. (2024). *Rl or not rl? parsing the processes that support human reward-based learning.* PsyArXiv.

Collins, A. G. (2018). The tortoise and the hare: Interactions between reinforcement learning and working memory. *Journal of cognitive neuroscience*, *30*(10), 1422–1432.

Collins, A. G., Brown, J. K., Gold, J. M., Waltz, J. A., & Frank, M. J. (2014). Working memory contributions to reinforcement learning impairments in schizophrenia. *Journal of Neuroscience*, *34*(41), 13747–13756.

Collins, A. G., & Frank, M. J. (2012). How much of reinforcement learning is working memory, not reinforcement learning? a behavioral, computational, and neurogenetic analysis. *European Journal of Neuroscience*, *35*(7), 1024–1035.

De Martino, B., & Cortese, A. (2023). Goals, usefulness and abstraction in value-based choice. *Trends in Cognitive Sciences*, *27*(1), 65–80.

Frömer, R., Dean Wolf, C. K., & Shenhav, A. (2019). Goal congruency dominates reward value in accounting for behavioral and neural correlates of value-based decision-making. *Nature communications*, *10*(1), 4926.

McDougle, S. D., Ballard, I. C., Baribault, B., Bishop, S. J., & Collins, A. G. (2022). Executive function assigns value to novel goal-congruent outcomes. *Cerebral Cortex*, *32*(1), 231–247.

Miller, K. J., Shenhav, A., & Ludvig, E. A. (2019). Habits without values. *Psychological review*, *126*(2), 292.

Molinaro, G., & Collins, A. G. (2023a). Human hacks and bugs in the recruitment of reward systems for goal achievement. In *Proceedings of the annual meeting of the cognitive science society* (Vol. 45).

Molinaro, G., & Collins, A. G. (2023b). Intrinsic rewards explain context-sensitive valuation in reinforcement learning. *PLoS Biology*, *21*(7), e3002201.

O'Reilly, R. C. (2020). Unraveling the mysteries of motivation. *Trends in cognitive sciences*, *24*(6), 425–434.

Palminteri, S., Justo, D., Jauffret, C., Pavlicek, B., Dauta, A., Delmaire, C., ... others (2012). Critical roles for anterior insula and dorsal striatum in punishment-based avoidance learning. *Neuron*, *76*(5), 998–1009.

Rigoux, L., Stephan, K. E., Friston, K. J., & Daunizeau, J. (2014). Bayesian model selection for group studies—revisited. *Neuroimage*, *84*, 971–985.

Senta, J., Bishop, S., & Collins, A. G. (2025). Dual process impairments in reinforcement learning and working memory systems underlie learning deficits in physiological anxiety. *bioRxiv*, 2025–02.

Wilson, R. C., & Collins, A. G. (2019). Ten simple rules for the computational modeling of behavioral data. *Elife*, *8*, e49547.